



ENVIRONMENTAL HEALTH PERSPECTIVES

<http://www.ehponline.org>

Computational Exposure Science: An Emerging Discipline to Support 21st-Century Risk Assessment

Peter P. Egeghy, Linda S. Sheldon, Kristin K. Isaacs,
Halûk Özkaynak, Michael-Rock Goldsmith, John F. Wambaugh,
Richard S. Judson, and Timothy J. Buckley

<http://dx.doi.org/10.1289/ehp.1509748>

Received: 21 January 2015

Accepted: 30 October 2015

Advance Publication: 6 November 2015

Note to readers with disabilities: *EHP* will provide a [508-conformant](#) version of this article upon final publication. If you require a 508-conformant version before then, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.



National Institute of
Environmental Health Sciences

Computational Exposure Science: An Emerging Discipline to Support 21st-Century Risk Assessment

Peter P. Egeghy,¹ Linda S. Sheldon,^{1§} Kristin K. Isaacs,¹ Halûk Özkaynak,^{1§} Michael-Rock Goldsmith,^{1‡} John F. Wambaugh,¹ Richard S. Judson,¹ and Timothy J. Buckley¹

¹U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, North Carolina, USA; [§]Retired; [‡]Current affiliation: Chemical Computing Group, Montreal, Quebec, Canada

Address correspondence to Peter P. Egeghy, Office of Research and Development, U.S. Environmental Protection Agency, 109 TW Alexander Dr., Mail Code E205-04, Research Triangle Park, NC 27711 USA. Telephone: (919) 541-4103. E-mail: egeghy.peter@epa.gov

Running title: Computational exposure science for risk assessment

Acknowledgments: We thank Dr. Jennifer Orme-Zavaleta, Director, U.S. EPA National Exposure Research Laboratory, and Dr. Kevin Crofton, Deputy Director, U.S. EPA National Center for Computational Toxicology, for many helpful comments and suggestions; Dr. Daniel Vallero, Senior Scientist, U.S. EPA National Exposure Research Laboratory, for help in conceptualizing “computational exposure science;” John Barton of SRA International for assistance with editing; and the anonymous reviewers for their time, effort, and engagement in improving this manuscript.

Disclaimer: The U.S. Environmental Protection Agency, through its Office of Research and Development, has provided administrative review of this article and approved it for publication.

Competing financial interests: The authors declare no actual or potential competing financial interests.

Abstract

Background: Computational exposure science represents a frontier of environmental science that is emerging and quickly evolving.

Objectives: In this commentary, we define this burgeoning discipline, describe a framework for implementation, and review some key ongoing research elements that are advancing the science with respect to exposure to chemicals in consumer products.

Discussion: The fundamental elements of computational exposure science include the development of reliable, computationally efficient predictive exposure models; identification, acquisition, and application of data to support and evaluate these models; and generation of improved methods for extrapolating across chemicals. We describe our efforts in each of these areas and provide examples that demonstrate both progress and potential.

Conclusions: Computational exposure science, linked with comparable efforts in toxicology, is ushering in a new era of risk assessment that greatly expands our ability to evaluate chemical safety and sustainability and to protect public health.

Introduction

Traditional, hazard-driven, single-chemical risk assessment practices that follow the 1983 National Research Council (NRC) paradigm (NRC 1983) cannot keep pace with the vast and growing numbers of chemicals in commerce (Harvey et al. 1995; Whittenberger et al. 1984). A well-defined, quantitative, and defensible means of identifying those with the greatest risk potential is needed (Judson et al. 2009; NRC 2007b, NRC 2009; Van Leeuwen et al. 2007), with exposure considerations providing a critical context for allocation of limited resources (Cohen Hubal 2009; Dellarco et al. 2010; Egeghy et al. 2011; Sheldon and Cohen Hubal 2009). Elevating the role of exposure science, however, will require development and application of efficient and reliable computational models that make full use of the rich and growing sources of accessible exposure-relevant information (NRC 2012). We propose a new discipline, called “computational exposure science,” that expands the knowledge and current methods used in the field of exposure assessment by bringing in novel data sources and new computational technologies.

A key driver of computational exposure science is the reinvigoration of interest in the significant role that the environment plays in disease etiology that has accompanied the conceptualization of the “exposome” (Prüss-Üstün and Corvalán 2006; Rappaport and Smith 2010; Wild 2005). The exposome refers to the totality of an individual’s environmental exposures from conception onwards. Integrating external agents, internal response, and the social, cultural, and ecological contexts of exposure, it was conceived to complement the genome for investigation of disease etiology (Wild, 2012). The rise in increasingly prevalent diseases, such as autism, asthma, and childhood leukemia, (Hertz-Picciotto and Delwiche 2009; Meeker 2012; Perrin et al. 2007) against a backdrop of widespread human exposure to industrial chemicals, as revealed by

biomonitoring surveys, (Becker et al. 2007; Centers for Disease Control and Prevention 2009; Park et al. 2012) confers a societal obligation to more comprehensively understand exposures. Traditional strategies for evaluating chemical exposures have left us without even the most basic information on exposures for the vast majority of chemicals in commerce (Egeghy et al. 2012; Muir and Howard 2006; Schwarzman and Wilson 2009), but a new era of systems thinking promises to transform exposure science.

Exposure science is responding to advances in technology (Cohen Hubal 2009). Rapid improvements in computing hardware and software have led to the emergence of efficient computational approaches to collect data, simulate complex processes, and systematically evaluate models. Simultaneously, diverse and “big” data are becoming increasingly available. New social-media-based methods of obtaining perception and behavior information are being developed (Eysenbach 2009), and further development of low-cost sensors soon will empower “citizen scientists” to provide a greater range of data, including chemical concentrations, using omnipresent technologies, such as smartphones (Dickinson et al. 2010; Snyder et al. 2013).

These public health drivers and novel scientific and technological advances are facilitating the development of computational exposure science as an emerging dimension of exposure science, akin to the emergence of computational toxicology over a decade earlier (el-Masri et al. 2002; Kavlock and Dix 2010). With various representations of computational exposure science beginning to appear in the published literature (Dionisio et al. 2015; Isaacs et al. 2014; Shin et al. 2015; Wambaugh et al. 2013), this commentary is intended to define the emerging discipline, establish a conceptual framework, and provide some illustrative examples of research being conducted to advance the field of exposure assessment using chemical ingredients of consumer products as an example.

Discussion

Defining Computational Exposure Science

We define computational exposure science as the integration of advances in chemistry, computer sciences, mathematics, statistics, and social and behavioral sciences with new and efficient models and data collection methods to reliably and effectively forecast real-world exposures to natural and anthropogenic chemicals in the environment. Computational exposure science aims to link exposures to health outcomes through the application of environmental informatics and advanced computational tools, as previously envisioned (Cohen Hubal 2009; Cohen Hubal et al. 2010; Sheldon and Cohen Hubal 2009), and to take full advantage of scientific innovation and the resulting abundance of newly available information for predictive, rapid, and high-throughput exposure assessment.

Although computational exposure science is conceptualized in the spirit of computational biology, it builds on a rich history of computational models for understanding environmental science that dates to the early 1900s (NRC 2007a). For key environmental chemicals, decades of observational studies have enabled the evolution of empirical and mechanistic models that reliably can explain the distribution and fate of chemicals in biological and environmental media. (Tornero-Velez et al. 2012; Zartarian et al. 2012). Multimedia fate models have greatly expanded modeling capacity to large inventories of chemicals. These models predict human exposure using mechanistic mass balance equations and food web bioaccumulation calculations to describe transfer between environmental compartments (e.g., air, water, soil, biota) (Arnot et al. 2012; Bennett et al. 2002; Diamond et al. 1994), but require few inputs other than easily obtainable physical/chemical descriptors. In the era of big data and informatics, analytical solutions are

giving way to more numerical, computational, and systems-focused methods. Nonetheless, a computational exposure science approach is not a replacement for more traditional monitoring, survey, and modeling methods used in exposure science. While computational exposure science provides the opportunity to investigate a more expansive range of chemicals, it also carries much greater levels of uncertainty. Risk context (NRC 2009) should determine whether rapid or highly refined approaches are employed. Moreover, the two approaches complement each other, as the screening-level results of the predictive models can identify targets for measurement, and the acquisition of new data through measurements and field studies is needed to evaluate and improve computational exposure science methods (NRC 2007b).

Applying Computational Exposure Science

Our ambitious goal for computational exposure science at the U.S. Environmental Protection Agency (EPA) is to rapidly and defensibly predict screening-level population exposure and intake dose rates for any existing or new chemical, even if little data exist beyond chemical structure. As depicted in Figure 1, understanding exposure to any chemical requires linkages from chemical functional role; to formulated component of consumer products; to identification of scenarios involving chemical release, media concentrations, and human contact; and, ultimately, to models estimating uptake and dose. The functional role of a chemical (i.e., how it is used in processes or products) is determined by its inherent chemical properties, which are imparted by chemical structure.

Predicting exposure and dose from chemical structure requires viable approaches for integrating information on the function of a chemical both with mechanistic fate and transport processes and with social and behavioral science descriptions of consumer product use and activities. Two

distinct realms of information, or “source data,” that are critical for estimating exposure are illustrated in Figure 1. These realms represent (1) inherent chemical, physical, and molecular properties and (2) decisions and other behaviors that govern product use. The individual components are arranged to visually indicate whether the data streams characterizing each component are mainly in the domain of “inherent chemical properties” or “human decisions and behavior” (beyond traditional exposure factors). As examples, media concentrations can be predicted using a multimedia mass balance model and a chemical release factor, both of which can be parameterized using the inherent properties of the chemical (Arnot et al., 2012). Product purchase and use, however, are the result of complex individual-level decisions that drive overall consumer behaviors in a stochastic fashion (Rand and Rust, 2011). The position of the “exposure” component reflects that the process is dependent as much on inherent properties as it is on complex social, psychological, and economic drivers.

The system of linkages depicted in Figure 1 may not seem like a radical departure from the traditional, linear source-to-dose continuum (Lioy 1990), nor is it intended as one. Instead, it expands on the modification proposed in the NRC *Exposure Science in the 21st Century* report (Lioy and Smith 2013; NRC 2012). Specifically, it expands the “upstream” factors of exposure to formulation science (i.e., functional role, product formulation) and to behavior informatics (i.e., product purchase, product use) and identifies the types of data streams required to parameterize each component. This addition of upstream processes provides a natural mechanism by which the product manufacturing stage of life cycle (i.e., product formulation) can be linked more directly to human behavior.

Application of computational exposure science to characterize each process in Figure 1 for thousands of chemicals will require one or more of the following generalizable activities:

- identify and acquire multiple data streams from traditional and nontraditional sources;
- extrapolate model input parameters from data-rich to data-limited chemicals;
- integrate multiple data streams for large sets of chemicals through rapid, efficient, and reliable models;
- evaluate model performance by systematically comparing model predictions with measured values obtained through targeted or non-targeted analyses and applying quantitative model sensitivity and uncertainty analyses to identify key data limitations and sources of uncertainty; and
- continuously acquire and incorporate new data streams that address key uncertainties and perform such refinements in an iterative and self-consistent manner.

An immediate challenge with computational exposure science is in identifying and integrating those data streams, particularly those outside of the traditional realm of exposure science, that are essential for understanding, parameterizing, and evaluating interactions between (chemical) stressors and (human) receptors. For example, commercial market research data and Internet search volume analytics remain largely unexploited for understanding consumer behaviors and their differences by region and demographics. Novel analytical tools (e.g., social network analysis, natural language processing) must be explored to facilitate the integration of nontraditional data streams into exposure assessment, just as ontologies are being developed to integrate exposure information with other disciplines (Goldsmith 2012; Mattingly et al. 2012).

To parameterize the processes in Figure 1 for the vast number of data-limited chemicals in commerce, it is necessary to extrapolate chemical-behavior patterns (fate, transport, intake, etc.) from relatively data-rich chemicals (often using quantitative structure-activity and relationship [QSAR] methodologies), with the explicit understanding that such extrapolation is fraught with

uncertainty and demands empirical evaluation (Molitor et al. 2007; Oreskes 1998; Wambaugh et al. 2013).

Innovative modeling approaches are needed to understand relationships among data sources of varying complexity and quality and exposure-related factors, processes, and monitoring data. These include machine-learning classification models, which already have proved well-suited for pharmacokinetic and hazard-related contexts (Freitas et al. 2015; Liu et al. 2015; Zang et al. 2013), and agent-based models, which provide a new opportunity to predict exposure-relevant behavior as a function of characteristics of individuals, their environments, and their interactions (Luke and Stamatakis 2012).

A clear understanding of domain of applicability (i.e., the set of conditions under which use of the model is scientifically defensible) is critical for reliable application of models, as is appropriately quantifying the precision of mathematical models, evaluating their predictive value, and characterizing associated uncertainties. Care must be taken to ensure that the models truly reflect their assumed underlying theoretical constructs, particularly when relying on big data (Lazer et al. 2014; Oreskes 1998). Conventional evaluations of model predictions against available personal measurement data, along with advances in computational implementations of statistical methods for model and data evaluation (Markov Chain Monte Carlo sampling for Bayesian inference, in particular), provide a path forward for such evaluation (Molitor et al. 2007; Wambaugh et al. 2013; Zartarian et al. 2012). The results of such evaluation will guide the acquisition and incorporation of additional data to address key uncertainties and further refine models (NRC, 2007a).

Current Research Activities and Examples

The application of computational exposure science as described above has led to a set of strategic research efforts by U.S. EPA to advance high-throughput exposure predictions. We provide examples below describing development and application of methods for assessing exposure to consumer product chemicals pertaining to (1) development of reliable, computationally efficient predictive models; (2) identification, acquisition, and analysis of data supporting high-throughput exposure model parameterization and model evaluation; and (3) extrapolation of available data to predict behaviors of large inventories of data-limited chemicals.

For proof of concept, our current focus has been on assessing exposures to chemical ingredients of consumer products under the construct that exposure to a chemical is a function of the type of product in which the chemical can be found and human activity patterns related to that product. The general strategy has been to identify products, map products to chemical ingredients, map products to use patterns and exposure scenarios, and then employ scenarios to model chemical exposures by route and pathway.

To supplement the National Library of Medicine's Household Products Database (NLM, 2015) for information on product ingredients, U.S. EPA has built the Consumer Product Chemical Profiles database (CPCPdb) (Goldsmith et al. 2014). Using optical character recognition and automated parsing to extract information from publicly available product material safety data sheets, CPCPdb has been populated with roughly 1800 unique chemicals in 353 product categories. EPA also has developed a database (Chemical Product Categorization database [CPCat]) of various levels of chemical use information for more than 40,000 chemicals (Dionisio

et al. 2015). CPCPdb has been consolidated into CPCat, and both databases are available through EPA's online warehouse of chemical data known as the Aggregated Computational Toxicology Resource (ACToR) (EPA, 2015). Use-related data within ACToR already have been shown to correlate with exposures inferred from biomonitoring (Wambaugh et al. 2014), and these databases provide a footing for the development of modeling systems to predict chemical functional use (based on properties), and then, from functional use, the types of products in which chemicals are likely to be found ("use profiles").

As QSARs are used to extrapolate physicochemical and pharmacokinetic properties across chemicals, similar models are being developed to determine relationships between predicted properties or structural descriptors and chemical functional role in products and to probabilistically predict weight fractions by functional role in consumer products for large lists of chemicals. Such analyses may eventually aid in identification of the underlying inherent chemical properties (molecule features) that confer the desired properties. As such, these computational exposure modeling methods can be repurposed to design safer ingredients or identify safer, readily available alternatives.

For the purposes of high-throughput exposure assessment, U.S. EPA has developed a new modeling approach, the Stochastic Human Exposure and Dose-Simulation-High-Throughput (SHEDS-HT) model (Isaacs et al. 2014), that combines use profiles with consumption information and then maps these factors to exposure scenarios. SHEDS-HT is based on the methods and algorithms of the SHEDS model for multimedia pollutants, commonly known as SHEDS Multimedia (Zartarian et al. 2012), but the fugacity, residential, and dietary modules have been reduced numerically and operationally to decrease user burden and increase run speed while maintaining critical features. A fugacity-based source-to-concentration module estimates

indoor concentrations by media (air, dust, and surfaces). Concentration estimates, along with relevant exposure factors and human activity data, then are used by the model to rapidly generate population-specific distributions of potential residential exposures via dermal, nondietary ingestion, and inhalation pathways. Estimated population dietary exposures are combined with the residential exposure predictions to produce total exposure estimates.

The development of SHEDS-HT, together with informatics-based methods of obtaining chemical use information, led to a significant increase in the speed of probabilistic exposure assessment and in the numbers of chemicals assessed. For example, the premier, higher-tier SHEDS Multimedia (SHEDS-MM) model had been applied to fewer than 10 chemicals over the past 15 years in support of high-priority regulatory support activities (Figure 2). In contrast, the first generation of SHEDS-HT extended the number to 15 by use of less detailed inputs and appropriate measurement surrogates. The addition of the simplified dietary module 1 year later increased the number to 330, and subsequent enhancement with information from consumer product ingredient databases brought the number of chemicals assessed to 2500 (Isaacs et al. 2014). We anticipate that current efforts toward the development of structure-to-function relationships will produce an accelerated rate of model parameterization that will enable screening-level forecasts for at least 10,000 chemicals by 2016 and perhaps twice that number in only 2 years later.

To assess new models such as SHEDS-HT and foster confidence for application in regulatory settings, we are developing techniques for evaluation of model predictions that are both systematic (e.g., to measure performance across a broad range of chemical stressors) and empirical (e.g., to determine how reliably available data support model estimates). We have relied on a well-defined framework using a Bayesian statistical methodology to draw inferences

from biomonitoring data for evaluating model predictions. This framework, called Systematic Empirical Evaluation of Models (Wambaugh et al. 2013, Wambaugh et al. 2014), provides (1) consensus exposure forecasts from multiple models and (2) an empirical determination of uncertainty in the resulting model predictions. This framework is a direct example of how traditional biomonitoring and exposure data can be used in concert with (perhaps less certain) computational exposure science modeling results.

Conclusions

The emerging discipline of computational exposure science represents an evolution of exposure science toward the identification of new data sources and application of innovative modeling techniques for understanding and quantifying human exposures to chemicals. The success of computational exposure science as a discipline will require that we also design and implement new research to collect the critical monitoring data needed to evaluate and improve the reliability of the next generation of models and to reduce the uncertainty in chemical exposure model predictions for screening and prioritization purposes and other applications (e.g., ecological impact analysis, life cycle analysis, broader sustainability evaluations). The non-targeted analysis of chemicals present in biological and environmental media performed using high-resolution mass spectrometry platforms will play a key role. The wealth of data produced by non-targeted measurement techniques can be used to generate and test hypotheses regarding the fate of chemicals as a function of, for example, their physical-chemical properties, use (applicative or functional), and source distance (near-field [applied to body or released indoors] versus far-field [released to outdoor environment]) but will require the innovative data analysis methods beyond those described herein. Results from both non-targeted analysis and computational models will be used to optimize future exposure data collection efforts. The symbiotic relationship between

methods, measurements, and modeling traditionally realized in exposure science is no less relevant within computational exposure science, but takes on a systems-focused and high-throughput form.

Although the examples provided here focus on EPA research, it should be acknowledged that other groups also are engaged in advancing computational exposure science. For example, the National Institutes of Health-funded Health and Exposome Research Center: Understanding Lifetime Exposures (HERCULES) (Hercules Exposome Research Center, 2015) and European Union-funded Health and Environment-wide Associations based on Large population Surveys (HEALS) (HEALS Consortium, 2015) projects each are bringing together novel technologies, data analysis techniques, and modeling tools to support exposome studies. Whereas our examples emphasize human exposure to consumer product chemicals, computational exposure science methods are also amenable for broader application, and similar approaches already are being evaluated for ecological receptors. As technology is rapidly evolving; the potential application of these methods will expand and the promise of minimizing significant adverse impacts of chemicals on human health becomes more attainable.

A critical mass of research around the themes of exposure modeling, statistics, and novel data streams is affirming computational exposure science as sufficiently distinct and mature to warrant description within the scientific literature. The emergence of computational exposure science has been motivated by both need and opportunity, in parallel with the earlier emergence of computational toxicology from toxicology. The availability of toxicity testing data for thousands of chemicals highlights the need for an exposure context to gauge risk and inform regulation (NRC 2007b; Thomas et al. 2013; Tice et al. 2013) and for exposure assessment to not be the rate-limiting step for high-throughput risk assessment. By identifying and defining this

new and rapidly emerging dimension of exposure science, we hope to foster its continued development in support of protection of health and the environment.

References

- Arnot JA, Brown TN, Wania F, Breivik K, McLachlan MS. 2012. Prioritizing chemicals and data requirements for screening-level exposure and risk assessment. *Environ Health Perspect* 120(11):1565-70.
- Becker K, Conrad A, Kirsch N, Kolossa-Gehring M, Schulz C, Seiwert M, et al. 2007. German environmental survey (GerES): Human biomonitoring as a tool to identify exposure pathways. *Int J Hyg Environ Health* 210:267-269.
- Bennett DH, Margni MD, McKone TE, Jolliet O. 2002. Intake fraction for multimedia pollutants: A tool for life cycle analysis and comparative risk assessment. *Risk Anal* 22:905-918.
- Centers for Disease Control and Prevention. 2009. Fourth national report on human exposure to environmental chemicals. Available: <http://www.cdc.gov/exposurereport/> [accessed 28 October 2015]
- Cohen Hubal EA. 2009. Biologically relevant exposure science for 21st century toxicity testing. *Toxicol Sci* 111:226-232.
- Cohen Hubal EA, Richard A, Aylward L, Edwards S, Gallagher J, Goldsmith M-R, et al. 2010. Advancing exposure characterization for chemical evaluation and risk assessment. *J Toxicol Environ Health B* 13:299-313.
- Dellarco V, Henry T, Sayre P, Seed J, Bradbury S. 2010. Meeting the common needs of a more effective and efficient testing and assessment paradigm for chemical risk management. *J Toxicol Environ Health B* 13:347-360.
- Diamond ML, Poulton DJ, Mackay D, Stride FA. 1994. Development of a mass balance model of the fate of 17 chemicals in the Bay of Quinte. *J Great Lakes Res* 20:643-666.

- Dickinson JL, Zuckerberg B, Bonter DN. 2010. Citizen science as an ecological research tool: Challenges and benefits. *Annu Rev Ecol Evol Syst* 41:149-172.
- Dionisio KL, Frame AM, Goldsmith M-R, Wambaugh JF, Liddell A, Cathey T, et al. 2015. Exploring consumer exposure pathways and patterns of use for chemicals in the environment. *Food Chem Toxicol* 2:228-237.
- Eghehy PP, Vallero DA, Cohen Hubal EA. 2011. Exposure-based prioritization of chemicals for risk assessment. *Environ Sci Policy* 14:950-964.
- Eghehy PP, Judson R, Gangwal S, Mosher S, Smith D, Vail J, et al. 2012. The exposure data landscape for manufactured chemicals. *Sci Total Environ* 414:159-166.
- EPA (U.S. Environmental Protection Agency). 2015. Aggregated Computational Toxicology Resource (ACToR). Available: <http://actor.epa.gov/> [accessed 28 October 2015].
- el-Masri HA, Mumtaz MM, Choudhary G, Cibulas W, De Rosa CT. 2002. Applications of computational toxicology methods at the Agency for Toxic Substances and Disease Registry. *Int J Hyg Environ Health* 205:63-69.
- Eysenbach G. 2009. Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *J Med Internet Res* 11:e11.
- Freitas AA, Limbu K, Ghafourian T. 2015. Predicting volume of distribution with decision tree-based regression methods using predicted tissue:plasma partition coefficients. *J Cheminform* 7:6.
- Goldsmith MR. 2012. Summary report for "Personal Chemical Exposure Informatics: Visualization and Exploratory Research in Simulations and Systems (PerCEIVERS)." EPA/600/R13/041. Environmental Protection Agency. Available:

<http://nepis.epa.gov/Exe/ZyPDF.cgi/P100LO5U.PDF?Dockey=P100LO5U.PDF> [accessed 28 October 2015].

Goldsmith MR, Grulke CM, Brooks RD, Transue TR, Tan YM, Frame A, et al. 2014.

Development of a consumer product ingredient database for chemical exposure screening and prioritization. *Food Chem Toxicol* 65:269-279.

Harvey T, Mahaffey KR, Velazquez S, Dourson M. 1995. Holistic risk assessment: An emerging process for environmental decisions. *Regul Toxicol Pharmacol* 22:110-117.

HEALS Consortium. 2015. Health and Environment-wide Associations based on Large population Surveys. Available: <http://www.heals-eu.eu/> [accessed 28 October 2015].

Hercules Exposome Research Center. 2015. Health and Exposome Research Center: Understanding Lifetime Exposures. Available: <http://emoryhercules.com/> [accessed 28 October 2015].

Hertz-Picciotto I, Delwiche L. 2009. The rise in autism and the role of age at diagnosis. *Epidemiology* 20:84-90.

Isaacs KK, Glen WG, Egeghy P, Goldsmith M-R, Smith L, Vallero D, et al. 2014. SHEDS-HT: An integrated probabilistic exposure model for prioritizing exposures to chemicals with near-field and dietary sources. *Environ Sci Technol* 48:12750–12759.

Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, et al. 2009. The toxicity data landscape for environmental chemicals. *Environ Health Perspect* 117(5):685-95.

Kavlock R, Dix D. 2010. Computational toxicology as implemented by the US EPA: Providing high throughput decision support tools for screening and assessing chemical exposure, hazard and risk. *J Toxicol Environ Health B* 13:197-217.

- Lazer DM, Kennedy R, King G, Vespignani A. 2014. The parable of google flu: Traps in big data analysis. *Science* 343(6176):1203-5.
- Lioy PJ. 1990. Exposure analysis and assessment for low-risk cancer agents. *Int J Epidemiol* 19 Suppl 1:S53-61.
- Lioy PJ, Smith KR. 2013. A discussion of exposure science in the 21st century: A vision and a strategy. *Environ Health Perspect* 121:405.
- Liu J, Mansouri K, Judson RS, Martin MT, Hong H, Chen M, et al. 2015. Predicting hepatotoxicity using toxcast in vitro bioactivity and chemical structure. *Chem Res Toxicol* 28:738-751.
- Luke DA, Stamatakis KA. 2012. Systems science methods in public health: Dynamics, networks, and agents. *Annu Rev Public Health* 33:357-376.
- Mattingly CJ, McKone TE, Callahan MA, Blake JA, Hubal EAC. 2012. Providing the missing link: The exposure science ontology ExO. *Environ Sci Technol* 46:3046-3053.
- Meeker JD. 2012. Exposure to environmental endocrine disruptors and child development. *Arch Pediatr Adolesc Med* 166:952-958.
- Molitor J, Jerrett M, Chang C-C, Molitor N-T, Gauderman J, Berhane K, et al. 2007. Assessing uncertainty in spatial exposure models for air pollution health effects assessment. *Environ Health Perspect* 115:1147.
- Muir DC, Howard PH. 2006. Are there other persistent organic pollutants? A challenge for environmental chemists. *Environ Sci Technol* 40:7157-7166.
- NLM (National Library of Medicine). 2015. Household Products Database. Available: <http://householdproducts.nlm.nih.gov/> [accessed 28 October 2015].

NRC (National Research Council). 1983. Risk Assessment in the Federal Government:

Managing the Process. Washington: National Academies Press. Available:

<http://www.nap.edu/catalog/366/risk-assessment-in-the-federal-government-managing-the-process> [accessed 28 October 2015].

NRC. 2007a. Models in Environmental Regulatory Decision Making. Washington: National

Academies Press. Available: <http://www.nap.edu/catalog/11972/models-in-environmental-regulatory-decision-making> [accessed 28 October 2015].

NRC. 2007b. Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington:

National Academies Press. Available: <http://www.nap.edu/catalog/11970/toxicity-testing-in-the-21st-century-a-vision-and-a> [accessed 28 October 2015].

NRC. 2009. Science and Decisions: Advancing Risk Assessment. Washington: National

Academies Press. Available: <http://www.nap.edu/catalog/12209/science-and-decisions-advancing-risk-assessment> [accessed 28 October 2015].

NRC. 2012. Exposure Science in the 21st Century: A vision and a Strategy. Washington:

National Academies Press. Available: <http://www.nap.edu/catalog/13507/exposure-science-in-the-21st-century-a-vision-and-a> [accessed 28 October 2015].

Oreskes N. 1998. Evaluation (not validation) of quantitative models. Environ Health Perspect

106 Suppl 6:1453-1460.

Park YH, Lee K, Soltow QA, Strobel FH, Brigham KL, Parker RE, et al. 2012. High-

performance metabolic profiling of plasma from seven mammalian species for simultaneous environmental chemical surveillance and bioeffect monitoring. Toxicology 295:47-55.

Perrin JM, Bloom SR, Gortmaker SL. 2007. The increase of childhood chronic conditions in the

United States. JAMA 297:2755-2759.

Prüss-Üstün A, Corvalán C. 2006. Preventing disease through healthy environments. Geneva:

World Health Organization. Available:

http://www.who.int/quantifying_ehimpacts/publications/preventingdisease/en/ [accessed 28 October 2015].

Rand W, Rust RT. 2011. Agent-based modeling in marketing: Guidelines for rigor. *Int J Res Mark* 28:181-193.

Rappaport SM, Smith MT. 2010. Epidemiology. Environment and disease risks. *Science* 330:460-461.

Schwarzman MR, Wilson MP. 2009. Science and regulation. New science for chemicals policy. *Science* 326:1065-1066.

Sheldon LS, Cohen Hubal EA. 2009. Exposure as part of a systems approach for assessing risk. *Environ Health Perspect* 117:1181-1184.

Shin H-M, Ernstoff A, Arnot JA, Wetmore BA, Csiszar SA, Fantke P, et al. 2015. Risk-based high-throughput chemical screening and prioritization using exposure models and in vitro bioactivity assays. *Environ Sci Technol* 49(11):6760-71.

Snyder EG, Watkins TH, Solomon PA, Thoma ED, Williams RW, Hagler GS, et al. 2013. The changing paradigm of air pollution monitoring. *Environ Sci Technol* 47:11369-11377.

Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, et al. 2013. Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework. *Toxicol Sci* 136:4-18.

Tice RR, Austin CP, Kavlock RJ, Bucher JR. 2013. Improving the human hazard characterization of chemicals: A Tox21 update. *Environ Health Perspect* 121:756-765.

- Tornero-Velez R, Davis J, Scollon EJ, Starr JM, Setzer RW, Goldsmith MR, et al. 2012. A pharmacokinetic model of cis- and trans-permethrin disposition in rats and humans with aggregate exposure application. *Toxicol Sci* 130:33-47.
- Van Leeuwen CJ, Patlewicz GY, Worth AP. 2007. Intelligent testing strategies. In: *Risk assessment of chemicals*. Springer, 467-509.
- Wambaugh JF, Setzer RW, Reif DM, Gangwal S, Mitchell-Blackwood J, Arnot JA, et al. 2013. High-throughput models for exposure-based chemical prioritization in the expocast project. *Environ Sci Technol* 47:8479-8488.
- Wambaugh JF, Wang A, Dionisio KL, Frame A, Egeghy P, Judson RS, et al. 2014. High throughput heuristics for prioritizing human exposure to environmental chemicals. *Environ Sci Technol* 8(21):12760-7.
- Whittenberger J, Bailar J, Doull J. 1984. *Toxicity testing strategies to determine needs and priorities*. Washington: National Academy Press. Available: <http://www.nap.edu/catalog/317/toxicity-testing-strategies-to-determine-needs-and-priorities> [accessed 28 October 2015].
- Wild CP. 2012. The exposome: from concept to utility. *Int J Epidemiol* 41(1):24-32.
- Wild CP. 2005. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 14:1847-1850.
- Zang Q, Rotroff DM, Judson RS. 2013. Binary classification of a large collection of environmental chemicals from estrogen receptor assays by quantitative structure-activity relationship and machine learning methods. *J Chem Inf Model* 53:3244-3261.

Zartarian V, Xue J, Glen G, Smith L, Tulse N, Tornero-Velez R. 2012. Quantifying children's aggregate (dietary and residential) exposure and dose to permethrin: Application and evaluation of EPA's probabilistic SHEDS-multimedia model. *J Expo Sci Environ Epidemiol* 22:267-273.

Figure Legends

Figure 1. Framework for Computational Exposure Science. The rounded rectangles represent components of computational exposure science required for high-throughput exposure assessment; their relative positions toward the left or the right represent the two generalized categories of source data (inherent chemical properties and human decisions, respectively).

Figure 2. The number of chemicals for which probabilistic exposure assessment has been performed has increased drastically over the past two years. Current efforts toward development of generic product formations and structure-to-function property-activity relationships will produce accelerated rates of evaluation.

Figure 1.

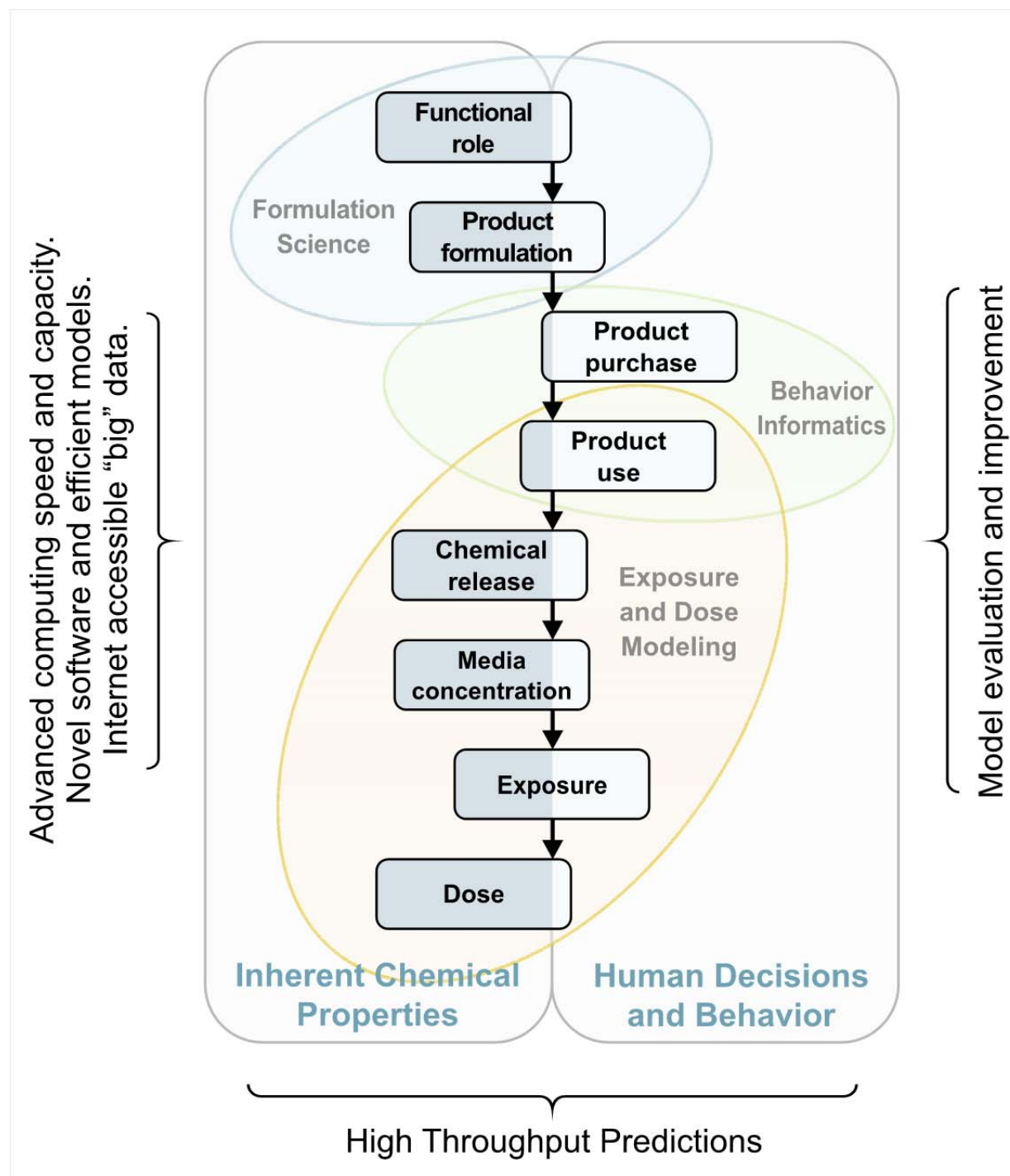


Figure 2.

